ISSN: 2821-157X **htpps://www.jasd.khadu.ir**

Journal of Aerospace Defense

Volume 3, Issue 4
Winter 2025
P.P 21-42

Research Paper;

A Novel Approach to Fake News Detection in Cyber Warfare Based on Transfer Learning and Stance Analysis

Mahmood Farokhian[1]

1- Department of computer engineering, Engineering Faculty, Arak University, Arak, Iran **E-mail:** farokhian@gmail.com

Article Information

Accepted:
2025/03/16

Received:
2024/12/27

Keywords:

Cybersecurity, Fake News, Deep Learning, Transfer Learning, Stance Detection

Corresponding Author:
Mahmood Farokhian

Email:
farokhian@gmail.com

Abstract

Background & Purpose: The rapid spread of fake news on social media has become a significant challenge in information security and cybersecurity, particularly in the context of passive defense. Early detection of such news is crucial for enhancing cybersecurity and controlling the dissemination of misinformation. This paper addresses the problem by leveraging the correlation between news headlines and their bodies to identify fake news, aiming to improve detection accuracy through advanced deep-learning techniques.

Methodology: This study proposes a novel approach based on transfer learning and stance analysis. Two independent BERT language models are fine-tuned separately on the headline and body of news articles. Deep neural networks analyze these components, and their correlation is measured to determine the authenticity of the news. The methodology involves preprocessing the text, encoding the headline and body using BERT, and employing a fully connected linear classifier to evaluate their alignment.

Findings: The results demonstrate that the proposed approach significantly improves detection accuracy compared to existing methods. By treating the headline and body as independent components and measuring their correlation, the model effectively identifies discrepancies indicative of fake news. The method achieves higher precision and recall, showcasing its potential for real-world applications in cybersecurity.

Conclusion: The study highlights the effectiveness of transfer learning and stance analysis in fake news detection. The proposed model offers a robust solution for identifying misinformation, contributing to enhanced cybersecurity and information integrity. Future work could explore multilingual datasets and additional contextual features to further refine the model's performance.

Mahmood Farokhian. A Novel Approach to Fake News Detection in Cyber Warfare Based on Transfer Learning and Stance Analysis. *Journal of Aerospace Defense*, Vol 3, No 4, 2025.

ISSN: 2821-157X | htpps://www.jasd.khadu.ir

فصل‌نامه علمی «دفاع هوافضایی»
دورهٔ ۳، شمارهٔ ۴
اسفند ۱۴۰۳
صفحات ۴۲-۲۱

مقاله پژوهشی

رویکردی نوین در تشخیص اخبار جعلی در جنگ سایبری مبتنی بر یادگیری انتقالی و تحلیل موضع

محمود فرخیان[۱]

۱. دکترای مهندسی نرم‌افزار کامپیوتر، گروه مهندسی کامپیوتر، دانشگاه اراک، اراک، ایران. **رایانامه:** farokhian@gmail.com

اطلاعات مقاله

تاریخ پذیرش:
۱۴۰۳/۱۲/۲۶

تاریخ دریافت:
۱۴۰۳/۱۰/۰۷

کلیدواژه‌ها:
امنیت سایبری، خبرجعلی، یادگیری عمیق، یادگیری انتقالی، تشخیص موضع

نویسنده مسئول:
محمود فرخیان
ایمیل:
farokhian@gmail.com

چکیده

انتشار سریع اخبار جعلی در شبکه‌های اجتماعی به یک چالش مهم در حوزه امنیت اطلاعات و امنیت سایبری، به‌ویژه در زمینه دفاع غیر فعال، تبدیل شده است. تشخیص زودهنگام این اخبار برای کنترل انتشار اطلاعات نادرست و ارتقای امنیت سایبری ضروری است. این مطالعه با استفاده از همبستگی بین تیتر و متن اخبار به شناسایی اخبار جعلی می‌پردازد و هدف آن بهبود دقت تشخیص از طریق تکنیک‌های پیشرفته یادگیری عمیق است.ما رویکردی نوین ترکیبی از یادگیری انتقالی و تحلیل موضع ارائه می‌کنیم. دو مدل مستقل زبانی BERT به‌طور جداگانه روی تیترها و متن اخبار تنظیم دقیق می‌شوند. شبکه‌های عصبی عمیق این اجزاء را تحلیل کرده و همبستگی آن‌ها را برای تعیین صحت خبر ارزیابی می‌کنند. روش‌شناسی شامل پیش‌پردازش متن، رمزگذاری تیتر و متن با استفاده از BERT و به‌کارگیری یک طبقه‌بند خطی کاملاً متصل برای ارزیابی همخوانی آن‌ها است. نتایج آزمایش‌ها نشان می‌دهد که رویکرد پیشنهادی دقت تشخیص را نسبت به روش‌های موجود به‌طور قابل توجهی بهبود می‌بخشد. با در نظر گرفتن تیتر و متن به‌عنوان اجزای مستقل و اندازه‌گیری همبستگی آن‌ها، مدل قادر است تناقض‌های نشان‌دهنده اخبار جعلی را به‌طور مؤثر شناسایی کند. این روش دقت و بازیابی بالاتری را ارائه می‌دهد و پتانسیل کاربرد در دنیای واقعی امنیت سایبری را نشان می‌دهد. این مطالعه اثربخشی یادگیری انتقالی و تحلیل موضع در تشخیص اخبار جعلی را نشان می‌دهد. مدل پیشنهادی ابزاری قوی برای شناسایی اطلاعات نادرست ارائه می‌دهد و به بهبود امنیت سایبری و یکپارچگی اطلاعات کمک می‌کند. پژوهش‌های آینده می‌توانند داده‌های چندزبانه و ویژگی‌های زمینه‌ای اضافی را برای ارتقای بیشتر عملکرد مدل بررسی کنند.

استناد: محمود فرخیان. رویکردی نوین در تشخیص اخبار جعلی در جنگ سایبری مبتنی بر یادگیری انتقالی و تحلیل موضع. مجله علمی پژوهشی دفاع هوافضایی، دوره ۳، شماره ۴، اسفند ۱۴۰۳.

۱– مقدمه

با رشد روزافزون رسانه‌های جمعی در سالهای اخیر تشخیص اخبار جعلی و حقیقی برای کاربران بسیار مشکل شده است. این اخبار می‌توانند به طور مستقیم بر آگاهی عمومی، تصمیم‌گیری‌های جامعه و اعتبار نهادها تأثیر بگذارند. اخبار جعلی اعتماد افراد جامعه به دولتها را کم کرده و میتواند باعث بروز بحران های خطرناکی در جامعه بشود. در زمینه جنگ سایبری، انتشار اخبار جعلی یکی از ابزارهای اصلی برای ایجاد اختلال در ساختارهای اجتماعی، کاهش اعتماد عمومی به منابع رسمی و تضعیف امنیت ملی است. این مسأله به‌عنوان بخشی از عملیات روانی و اطلاعاتی توسط گروه‌های مخرب یا دولت‌های متخاصم استفاده می‌شود. این موضوع به ویژه در زمان‌های بحرانی، نظیر انتخابات یا بحران‌های بهداشتی، اهمیت بیشتری پیدا می‌کند؛ چرا که اخبار جعلی می‌توانند به شایعات و نااطمینانی منجر شوند و اعتماد عمومی به منابع معتبر را کاهش دهند. به عنوان مثال در بحران غزه و اشغال آن توسط اسرائیل، شاهد بودیم که چگونه اخبار جعلی به‌عنوان یک ابزار جنگ اطلاعاتی استفاده می‌شد. این اخبار، گاه با هدف تضعیف روحیه مردم فلسطین و گاه برای ایجاد ابهام در جامعه جهانی در مورد واقعیت‌های این بحران منتشر می‌شدند. به‌عنوان مثال، اخبار نادرست درباره محل پناهگاه‌های ایمن یا روایت‌های جعلی از حوادث جنگی باعث گمراهی مردم و حتی تشدید بحران‌های انسانی شدند. چنین مواردی به وضوح نشان می‌دهند که اخبار جعلی می‌توانند به عنوان ابزاری مخرب در جنگ‌های سایبری و روانی استفاده شوند. در مثالی دیگر تأثیر قابل توجه انتشار اخبار جعلی بر رویدادهای مهم سال‌های اخیر یعنی انتخابات ریاست‌جمهوری آمریکا در ۲۰۱۶، همه‌پرسی برگزیت و همه‌گیری کووید ۱۹ هنوز خیلی دور از یادها نیست [۱].

رویکردهای مختلفی برای حل مساله اخبار جعلی توسط پژوهشگران در پیش گرفته شده است. دسته ای از روش‌ها از زمینه اجتماعی خبر برای بررسی صحت و سقم آن استفاده می‌کنند. این روش‌ها یا به بررسی الگوی انتشار اخبار جعلی و حقیقی در شبکه‌های اجتماعی و واکنش متفاوت کاربران شبکه‌ها به این دو دسته پرداخته یا از ویژگی‌هایی نظیر نویسنده خبر، منبع انتشار آن و یا اعتبار کاربران بازنشردهنده آن استفاده میکنند. دستهٔ دیگری از پژوهش‌ها نیز سعی در تشخیص خبر جعلی با ارزیابی محتوای آن دارند [۲–۴].

برخی از این پژوهش‌ها از روش‌های بر پایهٔ دانش برای بررسی اخبار جعلی استفاده میکنند؛ در واقع این کارها با خبر به مثابه یک یا چند گزاره منطقی برخورد می‌کنند و سعی میکنند میزان سازگاری این گزاره‌ها را با دانش قبلی ارزیابی کنند. دستهٔ دیگری از این پژوهش‌ها از تکنیکهای پردازش زبان طبیعی استفداده کرده و با مساله به عنوان یک مساله دسته‌بندی متن کلاسیک برخورد

می‌کنند. اما دستهٔ دیگری از روش‌ها سعی می‌کنند روش‌هایی ابداع کنند که سعی می‌کنند بر پایهٔ نظریاتی که در خصوص اخبار جعلی و دروغ در حوزه‌های روانشناسی، جرم‌شناسی و روزنامه‌نگاری وجود دارد، دست به تشخیص اخبار جعلی بزنند. به عنوان مثال گفتیم که متون دارای تخیل حاوی جزئیات و توصیفات عینی کمتری نسبت به متون دارای اعتبار هستند. بر همین اساس نسبت حسی‌بودن به عنوان یک معیار تشخیص متون جعلی بر اساس میزان جزییات متن پیشنهاد گردیده است [۵, ۶].

در سالهای اخیر توجه زیادی به استفاده از شبکه‌های عمیق در مسایل پردازش زبان طبیعی صورت گرفته است [۷, ۸]. یکی از مزیتهای مهم استفاده از شبکه عمیق برای پردازش متن، استخراج اتوماتیک ویژگی از متن است. در شبکه‌های عصبی نیاز به استخراج و کشف ویژگی به صورت دستی از متن نیست و لایه‌های ابتدایی شبکه خود عمل استخراج ویژگی‌های مناسب از متن را انجام می‌دهد. کارهایی که در این حوزه صورت گرفته است، نشان داده با پیش‌پردازش مناسب متن ورودی و تبدیل آن به عدد و خوراندن آن به شبکه‌های عصبی که مناسب پردازش متن طراحی شده‌اند می‌توان به نتایج برجسته‌ای در حل مسایل پردازش زبان طبیعی دست یافت. از آنجا که داده‌های متنی دارای ترتیب هستند بهتر است با آنها به عنوان دنباله‌داده‌های زمانی برخورد نمود. شبکه‌های عصبی عمیق کلاسیک مثل CNN، این نگاه زمانی به داده‌ها را ندارند و ترتیب داده‌ها برایشان مهم نیست [۹]. اما معماری‌هایی از شبکه‌های عصبی وجود دارد که مفهوم زمان موجود در داده‌ها را نیز در خروجی خود نهفته می‌کند. از جمله این دسته شبکه‌های عصبی عمیق می‌توان به RNN، GRU و LSTM اشاره کرد. یکی دیگر از معماری‌های پیشرفته شبکه‌های عصبی برای پردازش داده های ترتیبی، تبدیل‌کننده‌ها هستند که نتایج برجسته‌ای در تسک‌های پردازش زبان طبیعی ارایه داده‌اند. یکی از بهترین نتایج را شبکه‌های بازنمایی‌های رمزگذار دوطرفه از تبدیل‌کننده[1] در کاربردهای پردازش زبان طبیعی به دست آورده است که ما در این مقاله از آن استفاده می‌کنیم.

عموماً در خبرهای جعلی رابطهٔ استنباطی قوی‌ای بین عنوان خبر و متن آن وجود ندارد. بر این اساس، ما در این پژوهش سعی می‌کنیم تا با استفاده از یادگیری انتقالی، مدلی طراحی کنیم که بتواند موضع عنوان خبر و متن آن را نیست به هم تشخیص دهد. برای این کار، ابتدا عنوان خبر و مهم‌ترین بخش‌های متن خبر را انتخاب می‌کنیم. سپس، این داده‌ها را با استفاده از دو شبکه BERT موازی کدگذاری کرده و به یک طبقه‌بند خطی تمام‌متصل وارد کردیم. این شبکه خطی طوری آموزش

[1] Bidirectional Encoder Representations from Transformers (BERT)

می‌بیند که میزان همبستگی بین متن و عنوان خبر را که ملاک ما برای تشخیص اخبار جعلی است تخمین بزند. خروجی این شبکه عصبی مشخص‌کننده احتمال درست یا غلط بودن خبر است.

ادامه مقاله به این صورت سازمان‌دهی می‌شود. در بخش دوم مقاله به کارهای مشابهی که در زمینه تشخیص اخبار جعلی صورت گرفته است می‌پردازیم. در بخش سوم مفاهیم اساسی که در این پژوهش مورد اشاره قرار گرفته است را مرور میکنیم. در بخش چهارم مدل ارایه‌شده را معرفی کرده و در بخش پنجم نتایج به دست آمده از آزمایشات را بررسی می‌کنیم. در نهایت در بخش ششم به نتیجه‌گیری و ارایه کارهای پیشنهادی آینده می‌پردازیم.

۲- کارهای پیشین

سیستم‌های تشخیص اخبار جعلی را می‌توان بر اساس ورودی به سه دسته تقسیم کرد: روش‌های مبتنی بر اطلاعات بصری، روش‌های مبتنی بر زمینه اجتماعی و روش‌های مبتنی بر محتوا [۱۰]. به عنوان مثال در مقاله ی ارائه شده توسط وانگ و همکاران سعی شده است با استفاده از تصاویر همراه خبر و متن خبر و خوراندن آنها به شبکه‌های عمیق تمرین داده شده روی تصویر و متن و ترکیب اطلاعات به دست‌آمده از آنها موضع مناسبی در مقابل دروغ یا راست بودن خبر استخراج شود [۱۱].

پژوهش‌های مختلفی در زمینه تشخیص اخبار جعلی با استفاده از اطلاعات زمینه‌ای که خبر در آنها منتشر می‌شود صورت گرفته است که از جمله می‌توان به مطالعه ی وو و همکاران و شو همکاران اشاره کرد [۱۲, ۱۳]. به عنوان مثال ژو و همکاران (۲۰۱۹) دریافتند تفاوت معناداری در توزیع ساعات انتشار خبرهای جعلی و خبرهای درست در شبکه اجتماعی سیناویبو در طول شبانه‌روز قابل مشاهده شده است [۱۴]. وثوقی و همکاران مشاهده کردند خبرهای جعلی نسبت به خبرهای درست در بازهٔ گسترده‌تری توسط کاربران شبکه اجتماعی توییتر بازنشر می‌شوند [۴]. این تفاوت در گستردگی در چهار معیار عمق شارش (تعداد قدم ریتوییت از توییت اصلی)، اندازه شارش (تعداد کاربران مشارکت‌کننده) و حداکثر سطح (حداکثر تعداد مشارکت‌کننده در هر عمقی) قابل مشاهده است. آنها همچنین دریافتند الگوی انتشار اخبار جعلی در حوزه‌های مختلف با هم متفاوت است؛ به عنوان مثال اخبار جعلی سیاسی سریعتر از سایر حوزه‌ها در شبکه‌های اجتماعی منتشر می‌شوند. این تفاوت الگو در زبان‌ها، وب‌سایت‌ها و موضوعات مختلف نیز وجود دارد.

در روش‌های مبتنی بر محتوا تمرکز اصلی بر روی خود متن خبر می‌باشد. خود این روش‌ها به دو دسته اصلی تقسیم میشوند، روش‌های مبتنی بر دانش و روش‌های مبتنی بر سبک. روش‌های مبتنی بر دانش سعی بر ارزیابی خبر یا جمله بیان شده بر اساس روش واقعیت‌سنجی است. استنباط

درست یا غلط بودن خبر از طریق دانش در دو فاز صورت می‌پذیرد: در فاز اول که استخراج واقعیت نامیده می‌شود، دانش اغلب از صفحات وب استخراج گردیده و در فاز دوم یعنی واقعیت‌سنجی دانش استخراج شده از خبر مورد بررسی با دانش پایه که در پایگاه دانش گردآوری شده مطابقت داده شده تا صحت و سقم آن خبر مشخص گردد.

تورن و همکاران یک خط لوله طراحی کردند که در سه فاز به واقعیت‌سنجی ادعای مطرح شده با استفاده از جعبه‌های اطلاعات درون ویکی‌پدیا می‌پردازد [۶]. در فاز اول که بازیابی اسناد است، k نزدیکترین سند به ادعای مطرح شده بازیابی شد. در فاز دوم از میان k سند بازیابی شده، تعداد مشخصی جمله که بیشترین ربط را با ادعای مطرح شده دارند، انتخاب شد و در فاز سوم ادعای مطرح شده نسبت سه‌تایی‌های استخراج شده از جملات منتخب با روش شناسایی استلزام متنی سنجیده شد.

پن و همکاران به تشخیص اخبار جعلی با استفاده از گراف دانش پرداختند. آنها ابتدا از یک گراف دانش به عنوان دانش معتبر پایه استفاده کردند و یک مدل transE ساختند. یک مدل transE نیز برای مقاله خبری مورد بررسی ساخته شدو سپس با محاسبهٔ یک تابع انحراف نسبت به مدل transE پایه، مقاله در یکی از دو دسته جعلی یا معتبر دسته‌بندی شد [۱۵].

روش‌های مبتنی بر سبک سعی بر استخراج تعدادی ویژگی از متن نمود و با استفاده از این ویژگی‌ها دست به پیش‌بینی اعتبار خبر زدند. این مطالعات، بر اساس تحلیل و بررسی متن مقاله، به تشخیص جعلی بودن آن پرداخت. این تحلیل و بررسی در چهار سطح لغت، نحو، معنا و گفتمان صورت گرفت. در کاری که توسط اوت و همکاران و ابولینین انجام شد، از مدل‌های n-gram و خصایص آماری کلمات و حتی حروف موجود در متن به عنوان خصیصه‌های متن استفاده شد [۱۶]. در [۱۶–۱۸] از گرامرهای مستقل از متن احتمالی و POS استفاده شده است. همچنین کریمی و همکاران سعی کردند ابتدا با استفاده از تحلیل وابستگی بین جملات به عنوان واحدهای بلاغی نحوه وابستگی بین آنها را کشف کنند و سپس با تحلیل نحوه این وابستگی و میزان آن به تشخیص جعلی بودن یا نبودن آن بپردازند [۵]. در بسیاری از کارهای صورت گرفته، از مدل‌های یادگیری کلاسیک مختلفی از جمله درخت تصمیم‌گیری، SVM و KNN استفاده شده است. اما چالش اصلی در این پژوهش‌ها انتخاب ویژگی‌های مناسب از متن است. هیچ تضمینی وجود ندارد فریب نهفته شده در اخبار در ویژگی‌های مورد استفاده در هر یک از مدل‌ها را نمایان کند. استفاده از شبکه‌های عصبی عمیق این مزیت را دارد که طی پروسه تمرین بهترین ویژگی‌های نمایان کننده درست و غطط بودن متن را می‌تواند به طور خودکار استخراج کند [۱۹]. در کار ارائه شده توسط مونتی و همکاران از شبکه‌های عصبی هندسی و اعمال آنها روی مدل انتشار اخبار در شبکهٔ توییتر استفاده شد [۸]. شبکه‌های عصبی هندسی که در اصل از شبکه‌های عصبی پیچشی مشتق شده است نوعی شبکه

عصبی غیر اقلیدسی است که در تحلیل داده‌های گراف و مانیفولد کاربرد دارد. در پژوهشی دیگر، القمدی و همکاران به بررسی تأثیر روابط معنایی بین محتوای اخبار و تیترها و نظرات کاربران در تشخیص اخبار جعلی پرداختند [۲۰]. آن‌ها رویکردی نوآورانه با استفاده از تکنیک‌های یادگیری عمیق، از جمله مدل BERT و شبکه‌های توجه چندبخشی[1] را پیشنهاد کردند.

۳- مفاهیم پایه

۳-۱- یادگیری انتقالی

برای استفاده شبکه‌های عصبی عمیق در تشخیص اخبار جعلی اما یک چالش اساسی وجود دارد؛ یادگیری عمیق برای تمرین داده شدن نیاز به حجم عظیمی از داده‌ها دارد اما دسترسی ما به اخبار جعلی بسیار محدود است. نگاهی به مجموعه داده‌های خبر جعلی که تا کنون ساخته شده است نشان می‌دهد اکثر آنها اندازه‌هایی در حدود چندصد یا چند هزار دارند که برای تمرین دادن یک شبکه عصبی پیچیده با تعداد زیادی پارامتر مناسب نیست [۲۱]. در سالهای اخیر برای رفع این مشکل به یادگیری انتقالی روی آورده شده است. ایده اصلی یادگیری انتقالی این است که بتوان دانش به‌دست آمده در یک دامنه را در دامن دیگری به کار برد. برای این کار مدل شبکه عصبی ابتدا روی حجم عظیمی از داده‌های عمومی تمرین داده می‌شود و بعد مدل از پیش تمرین داده شده روی داده‌های مخصوص یک تسک خاص ریزتنظیم[2] شود [۲۲]. این روش یادگیری از سال‌ها پیش در رویه‌های پردازش تصویر مرسوم بوده است و به تازگی در پردازش زبان طبیعی هم مورد توجه قرار گرفته است. این ایده خصوصاً در تشخیص اخبار جعلی بسیار گره‌گشاست، زیرا مدل‌های یادگیری عمیق نیازمند حجم زیادی از داده‌ها هستند و عموماً جمع‌آوری اخبار جعلی و ساخت مجموعه داده‌های حجیمی که مناسب کاربرد در یادگیری عمیق باشد، کاری سخت و تا حدی نشدنی است.

برای یادگیری انتقالی در پردازش زبان طبیعی پیش‌تمرین به منظور ساخت یک مدل زبانی پایه مورد استفاده قرار می‌گیرد. کار مدل‌های زبانی در حالت کلی پیش‌بینی کلمه، یا جمله جا افتاده در متن است. مدل‌های زبانی دو خاصیت مهم دارند که آنها را کاندیدای خوبی برای استفاده به عنوان مدل پایه میکند. اول اینکه می‌توانند بر روی داده‌های بدون برچسب آموزش ببینند و بنابراین حجم عظیمی از داده‌ها برای یک یادگیری قوی در اختیار آنان هست. دوم این که یک مدل زبانی که به خوبی از پیش تمرین داده شده است، می‌تواند فهم خوبی از ساختار و معنای زبان به ما بدهد که این

---

[1] Multimodal
[2] Fine Tune

فهم در تسکهای داون استریم بسیار مورد نیاز و کمک کننده است.

مدل زبانی که در این فرآیند ساخته می‌شوند در واقع جایگزین تعبیه‌های کلمه ایستا مانند Word2Vec و Glove می‌شوند [۲۳, ۲۴]. این مدل زبانی را می‌توان نوعی تعبیه کلمه پویا نامید. تعبیه‌های کلمه ایستا مستقل از متن بودند. یک کلمه وقتی در متن‌های های مختلف به کار برود می تواند معانی مختلفی بدهد، اما در تعبیه ایستا این کلمه همیشه به یک بردار ثابت نگاشت می‌شد. برای مثال کلمه run در دو جمله "Robert is running a club" و " Robert is running in a marathon" دو معنی کاملا متفاوت دارند اما نمایش هر دوی آنها در word2vec یکسان است. در مدل‌های زبان که بر اساس یادگیری انتقالی ساخته می‌شوند، این کل متن است که به دنباله‌ای از بردارهای تعبیه شده نگاشت می‌شود، نه یک کلمه. بنابراین نمایش هر کلمه توسط یک بردار با توجه به متن صورت می‌پذیرد.

برای انتقال یادگیری در پردازش زبان طبیعی از معماری‌های شبکه عصبی مختلفی استفاده می شود. برای مثال، ELMo از یک LSTM دو جهته استفاده می‌کند [۲۵]. اما معمول‌ترین معماری مورد استفاده تبدیل‌کننده‌ها هستند. GPT2 و BERT دو مدل زبانی جدید هستند که هر دو از تبدیل‌کننده‌ها برای انتقال یادگیری در پردازش زبان طبیعی استفاده می کنند و در کاربرد به نتایج درخشانی دست یافته‌اند [۲۶].

۳-۲- تبدیل‌کننده‌ها

اخیرا، اکثر سیستم‌های یادگیری از شبکه‌های عصبی بازگشتی دروازه‌داری[1] مانند LSTM و GRU برای پردازش زبان طبیعی استفاده می‌کردند که برای پردازش داده‌های متوالی طراحی شده بودند. این معماری ها پس از معرفی مکانیسم توجه، لایه های توجه را نیز به معماری خود اضافه کردند تا ارتباط بین اجزای توالی را بهتر یاد بگیرند و اجزا را با توجه به اهمیت آنها در مدل وزن کنند [۲۷]. در سال های اخیر با توجه به نتایج عالی تبدیل‌کننده‌ها در کارهای پردازش زبان طبیعی، در اکثر تحقیقات، تبدیل‌کننده‌ها جایگزین معماری‌های قبلی شده‌اند. تبدیل‌کننده‌ها تماماً بر اساس مکانیسم توجه ساخته می شوند زیرا این واقعیت وجود دارد که مکانیزم توجه به تنهایی قادر به برآوردن عملکرد یادگیری که از RNN برمی‌آید، می‌باشد [۴].

یکی از آخرین و امیدوارکننده‌ترین مدل‌های تبدیل‌کننده، BERT است. BERT که توسط متخصصان گوگل ارایه شده روی حجم عظیمی از داده‌های بدون برچسب با توجه به دو هدف

[1] Gated Recurrent Neural Network

پیش‌تمرین داده شده است:

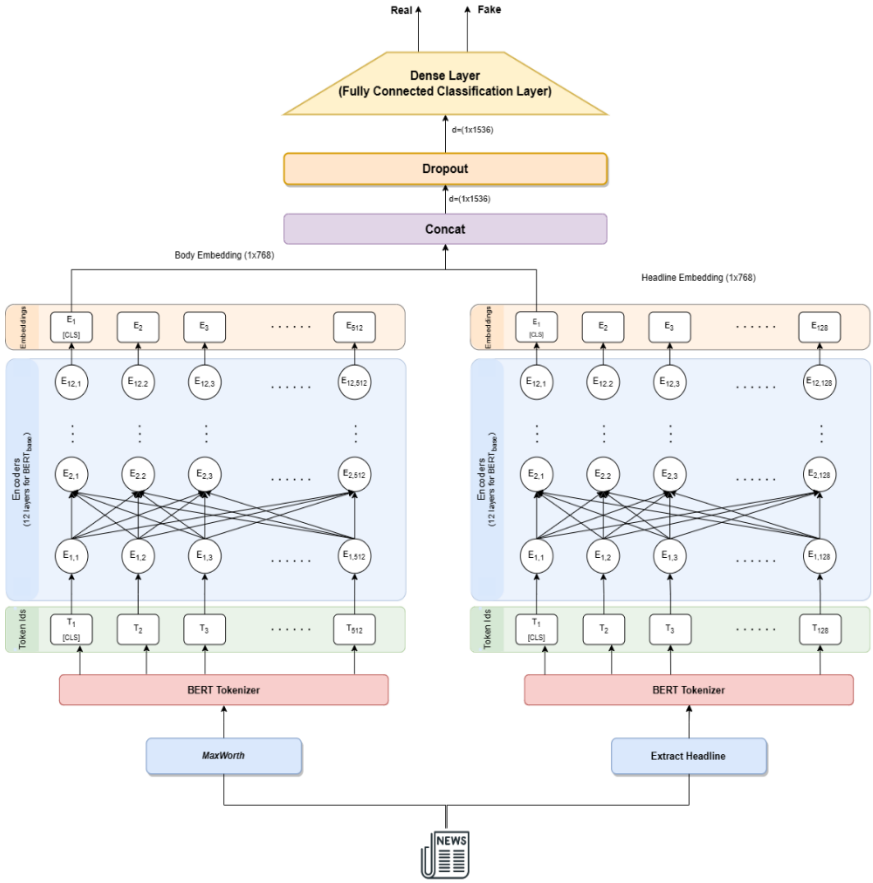
مدلسازی زبان مسک شده: به این منظور حدود ۱۵ درصد از کلمات جمله مسک میشوند و مدل باید بتواند با استفاده از زمینه کلمات آنها را پیش‌بینی کند. باید در نظر داشت بر خلاف مدل‌های قبلی تبدیل‌کننده مثل GPT، برت دوطرفه است، یعنی از محتوای پیش و پس از کلمه برای حدس زدن آن استفاده می‌کند. نشان داده شده است که این امر میتواند باعث افزایش کارآیی مدل نسبت به مدل‌های یکطرفه قبلی مثل GPT در تسکهای سطح جمله از جمله سوال–جواب (SQuAD 1.1) و استناج در زبان طبیعی (MNLI) شود [۲۸, ۲۹]. همچنین وابستگی مدل به جهت زبان را از بین می‌برد و برای زبانهای راست به چپ نیز به اندازه زبان‌های چپ به راست کارآیی دارد.

پیش‌بینی جمله بعدی: با دادن دو جمله مدل باید حدس بزند جمله دوم بعد از جمله اول قرار دارد یا نه؟

برای استفاده برت در تسکهای پایین‌دست کافی است با تمرین دادن روی داده‌های اختصاصی تسک تنظیم دقیق شود. برای بهره‌گیری از مدل برت از پیش‌تمرین داده شده در تسکهای پایین دست همان معماری برت استفاده می‌شود و فقط لایه‌های خروجی را متناسب با خواسته تسک مربوطه به آن اضافه میکنیم. در زمان تنظیم دقیق مدل با پارامترهای به دست آمده از پیش تمرین شروع می‌کنیم و تمام پارامترها در زمان تنظیم دقیق، تنظیم می‌شوند.

۴– راهکار پیشنهادی

طرحی از مدل پیشنهادی ما که ما آن را StanceBert نامیده‌ایم در شکل ۱ نشان داده شده است. از یک سو تیتر خبر با حداکثر طول ۱۲۸ وارد یک شبکه BERT می شود. از طرف دیگر با استفاده از الگوریتم MaxWorth، مناسب ترین بازه متن خبری از بقیه متن خبر انتخاب می شود. متن بازهٔ انتخاب شده نیز وارد یک شبکه BERT مجزا با حداکثر طول ورودی ۵۱۲ می‌شود. خروجی هر دو شبکه، که معنای معنایی تیتر و متن خبر را نشان می دهد، وارد یک شبکه عصبی چندلایهٔ با اتصال کامل می‌شود. وظیفه این لایه مقایسه نمایش تیتر و بدنه خبر با یکدیگر است. در اخبار واقعی بدنه هماهنگی زیادی بین تیتر و بدنه خبر وجود دارد اما در اخبار جعلی این‌طور نیست. در نهایت یک لایهٔ دراپ آوت نیز برای جلوگیری از بیش‌برازش به انتهای ساختار شبکه اضافه شده است.



شکل ۱ - نمای کلی مدل StanceBert

اندازه ورودی و خروجی هر لایه شبکه در جدول ۱ قابل مشاهده است.

جدول ۱ - اندازه ورودی و خروجی لایه های مدل

لایه	اندازه ورودی	اندازه خروجی
برت (متن خبر)	۵۱۲	۷۶۸
برت (تیتر)	۱۲۸	۷۶۸
الحاق	۲×۷۶۸	۱۵۳۶
حذف تصادفی	۱۵۳۶	۱۵۳۶
لایهٔ خطی	۱۵۳۶	۲

٤-۱- لایهٔ شبکهٔ BERT

معماری‌های مختلفی از برت وجود دارد که مهم‌ترین آنها را در جدول ۲ می‌بینید. در ارزیابی‌های این مدل‌ها دیده شده که مدل‌های uncased برای زبان انگلیسی اندکی بهتر کار می‌کنند. با توجه به اندازه نسبتاً کوچک مجموعه داده مورد استفاده ما استفاده از مدل لارج در تجربیات ما همان طور که

در [۲۶] اشاره شده، منجر به نتایج خوبی نشد. لذا در ادامه کار از مدل پایهٔ برت uncased استفاده کرده‌ایم.

جدول ۲ - معماری های از پیش تمرین داده شده BERT

مدل	تعداد لایه	تعداد نرون در هر لایه	حساسیت به حروف بزرگ و کوچک	تعداد پارامترها
$BERT_{large}$ -cased	۲۴	۱۰۲۴	بله	۳۴۰ میلیون
$BERT_{large}$ -uncased	۲۴	۱۰۲۴	خیر	۳۴۰ میلیون
$BERT_{base}$ –cased	۱۲	۷۶۸	بله	۱۱۰ میلیون
$BERT_{base}$ –uncased	۱۲	۷۶۸	خیر	۱۱۰ میلیون
$BERT_{base}$- Multilingual -cased	۱۲	۷۶۸	دارد	۱۱۰ میلیون
$BERT_{base}$- Multilingual –uncased	۱۲	۷۶۸	دارد	۱۱۰ میلیون
$BERT_{base}$ –Chinese	۱۲	۷۶۸	-	۱۱۰ میلیون

٤-۲- شبکهٔ خروجی

برای وظایف در سطح توکن مانند تگ کردن دنباله یا سوال جواب نمایش برداری تک توکن‌ها وارد شبکه خروجی میشود اما برای دسته‌بندی فقط نمایش توکن خاص اول متن (CLS) به عنوان نمایشی از کل متن وارد لایه خروجی می‌شود. شبکهٔ خروجی یک شبکه ساده چندلایه با اتصال کامل است و یک لایه دراپ آوت برای جلوگیری از اوورفیت شدن شبکه است. افزودن هرگونه پیچیدگی بیشتر به خروجی نه تنها باعث افزایش کارآیی شبکه نمی‌شود، بلکه اثر مکانیزم سلف اتنشن موجود در برت را نیز از بین می‌برد. آزمایش‌های بخش بعدی صحت این ادعا را نشان می‌دهند.

٤-۳- مجموعه داده

ما برای ارزیابی روش پیشنهادی خود از مجموعه داده Fakenewsnet استفاده کرده‌ایم. Fakenewsnet توسط شو و همکاران گردآوری شده است [۳۰]. فیک‌نیوزنت بزرگترین مجموعه داده اخبار کامل موجود است که حاوی ۲۲۶۱۶ خبر کامل گردآوری شده از سایتهای politifact.com و gossipcop.com است. اغلب مجموعه داده‌های اخبار جعلی از روی توییت‌ها، سرتیترها[1]، کامنت‌ها و ادعاها گردآوری شده که طول کوتاهی دارند و اساساً خبر به حساب نمی‌آیند. اما وب‌سایتهای مورد استفاده در این مجموعه داده حاوی خبر کامل هستند که گاهی طول آن به چندین صفحه می‌کشد.

[1] Headlines

علاوه بر این، اخبار موجود در این مجموعه داده دارای کیفیت مناسبی هستند و واقعاً خبرهایی بوده‌اند که در مقطعی مورد توجه بسیاری از کاربران قرار گرفته است. حتی خبرهای صحیح موجود در مجموعه داده، خبرهایی بوده‌اند که مشکوک به جعلی بودن، بوده‌اند و به همین خاطر در سایتهای gossipCop و politifact که کارشان بررسی اخبار جنجالی و بحث برانگیز است مورد اشاره قرار گرفته‌اند. آمار خبرهای صحیح و غلط و منابع خبری مورد استفاده شده در این مجموعه داده را به ما می‌دهد.

جدول ۳ - آمار اخبار مجموعه داده FakeNwesNet

منبع	دروغ	صحیح
PolitiFact.com	۴۲۰	۵۲۸
GossipCop.com	۴۹۷۴	۱۶۶۹۴

۵ –آزمایش‌ها و نتایج

ما مدل پیشنهادی خود را با سه مدل برجسته مقایسه کردیم که اخیراً در مقالات معتبر ظاهر شده‌اند و در ارزیابی‌های خود به نتایج خوبی دست یافته‌اند. اولین مدل شبکه توجه سلسله مراتبی سه سطحی (3HAN[1]) است [۳۱]. این مدل از GloVE به عنوان لایه تعبیه کلمات و سه لایه شبکه عصبی بازگشتی دروازه‌دار دو طرفه مجهز به مکانیزم توجه، به ترتیب به عنوان رمزگذار کلمه، رمزگذار جملات و رمزگذار سرخط خبر استفاده میکند. مدل دیگر هم مدل FakeBERT است که توسط کالیار و همکاران ارائه شده است و از تعبیه کلمات پویای BERT استفاده می‌کند [۳۲]. در این مدل خروجی لایهٔ BERT وارد یک شبکه خروجی پیچیده میشود که شامل چندین لایه شامل سه لایه شبکه عصبی پیچشی[2] موازی، دو لایه بیشینهٔ تجمع[3] و چندین لایه تنظیم کننده دیگر است.در نهایت مدل سوم یک مدل پایه CNN با استفاده از تعبیه GloVE است که از مقاله اصلی FakeNewsNet الهام گرفته شده است [۳۰].

هدف نهایی هر کلاسه‌بند دستیابی به بالاترین دقت ممکن است که به صورت نسبت تعداد حدسهای درست به کل تعداد نمونه‌ها است. اما از آنجا که معیار دقت به تنهایی به‌خصوص در کلاسهای نامتقارت میتواند ما را به اشتباه بیندازد، معیارهای صحتِ، یادآوری و امتیاز F1 را نیز برای

---

1 3 Layered Hirarchical AttentionNetwork
2 Convolutional Neural Network
3 Max Pooling

هر مدل اندازه گرفته‌ایم. همچنین یکی از معیارهای معتبر که میزان یادگیری مدل را نشان می‌دهد نمودار ROC و عدد RUC است که نشان دهنده سطح زیر نمودار ROC میباشد. این معیار را نیز برای همه مدلها محاسبه نموده‌ایم.

برای سنجش مدل‌ها، ما مجموعه داده را به دو بخش تقسیم کردیم: ۸۰ درصد برای آموزش و بقیه برای آزمون. سپس، مدل‌ها را با داده‌های آموزشی، آموزش دادیم و عملکرد مدل آموزش‌دیده را روی مجموعه آزمون اندازه‌گیری کردیم. در هر دوره آموزش نیز، ده درصد از داده‌های آموزش برای اعتبارسنجی درنظر گرفته می‌شد و باقی داده‌ها در آموزش مورد استفاده قرار می‌گرفت. فراپارامترهای پروسه آموزش را می توان در جدول ۳ مشاهده کرد. نرخ یادگیری اولیه در مقاله مرجع BERT ([۲۶]) بین $۵ \times e^{-5}$ تا e^{-5} توصیه شده و ما هم در اکثر آزمایش‌های خود $۲ \times e^{-5}$ را مناسب‌ترین نرخ یادگیری اولیه دیدیم. البته با توجه به اینکه بهینه‌ساز مورد استفاده AdamW است، این نرخ در طی فرآیند آموزش به طور مداوم تغییر می‌کند. AdamW یک الگوریتم بهینه سازی برای شبکه های عصبی است که از الگوریتم پایهٔ Adam (تخمین لحظه تطبیقی) به عنوان الگوریتم بهینه سازی اولیه استفاده می کند و با لحاظ‌کردن عنصر محوشدگی وزن ، جریمه ای بر وزن‌ها اعمال می کند تا از بیش‌برازش جلوگیری کند [۳۳].

جدول ۳- فراپارامترهای آموزش مدل

پارامتر	مقدار
نرخ یادگیری اولیه	۲×۱۰^{-۵}
اندازه دسته	۸
بهینه‌ساز	AdamW
اپسیلون (بهینه‌ساز)	۱×۱۰^{-۵}
احتمال حذف تصادفی	۰.۲
تابع خطا	Cross Entropy

نتایج مدل‌ها را می توان در جدول ۴ مشاهده کرد. همانطور که مشاهده می شود StanceBert بهترین نتیجه را در اکثر معیارهای عملکرد به دست آورده است.

جدول ٤ - معیارهای عمل‌کرد مدل‌ها

مدل	صحت	دقت	یادآوری	امتیاز F1	AUC
CNN	۰.۷۱۸	۰.۵۶۳	۰.۶۲۳	۰.۵۹۲	۰.۸۰۸
3HAN	۰.۸۲۹	۰.۶۸۸	۰.۴۹۶	۰.۵۷۶	۰.۸۲۵
FakeBERT	۰.۸۰۱	۰.۵۶۰	۰.۷۰۵	۰.۶۲۴	۰.۸۵۴
StanceBert	۰.۸۵۴	۰.۷۵۶	۰.۵۵۵	۰.۶۴۰	۰.۸۶۳

متوسط زمان اجرای آموزش هر مدل و تعداد دوره‌های مورد نیاز تا رسیدن به حداقل خطای اعتبارسنجی (طی اجراهای مختلف) در جدول ۵ آمده است. تعداد دوره[1] مورد نیاز، تعداد دوره‌ای است که بعد از آن مدل شروع به بیش‌برازش می‌کند و خطای اعتبارسنجی بالا می‌رود که این تعداد در اجراهای متفاوت ممکن است کمی متفاوت باشد. با توجه به اینکه مدل ما از دو شبکه برت بهره می‌برد زمان آموزش طولانی‌تر آن قابل انتظار بود.

جدول ۵ - زمان آموزش مدل‌ها

مدل	زمان آموزش	تعداد دوره مورد نیاز (متوسط)
StanceBert	۱۵۰ دقیقه	۲.۲ دور
FakeBERT	۹۵ دقیقه	۲.۹ دور
3HAN	۶۰ دقیقه	۲۷ دور
CNN	۴۳ دقیقه	۲۱ دور

۵-۱- ارزیابی انتخاب‌های طراحی

علاوه بر مقایسه مدل پیشنهادی خود با مدل‌های دیگر، ما قصد داشتیم به سوالاتی در رابطه با انتخاب‌هایی که در طراحی مدل خود انجام داده‌ایم، پاسخ دهیم. این سوالات به شرح زیر بود:

- آیا استفاده از دو BERT موازی باعث بهبود عملکرد مدل می شود؟
- آیا استفاده از الگوریتم MaxWorth به جای روش‌های سنتی خلاصه‌سازی متن منطقی است؟
- آیا پیچیدگی شبکه خروجی بر عملکرد مدل تأثیر می‌گذارد؟
- آیا طول متن ورودی بر عملکرد مدل تأثیر می گذارد؟

[1] Epoch

– آیا استفاده از مدل برت بزرگتر[1] کارآیی را افزایش نمی‌دهد؟

برای پاسخ به این سوالات آزمایش‌هایی انجام دادیم تا بتوان به کمک آنها انتخاب‌های طراحی مدل را ارزیابی کرد. نتیجه این آزمایش‌ها در ادامه این بخش ارائه شده است. برای معتبر بودن مقایسه‌ها، در هر مدل فقط جنبه مورد اشاره تغییر کرده و مدل در سایر جنبه‌ها دقیقاً شبیه مدل پیشنهادی اصلی ماست.

۵-۱-۱- روش انتخاب گسترهٔ متن

در مدل پیشنهادی ما برای انتخاب قسمت مناسب از متن خبر برای ورود به BERT، به جای استفاده از روش‌های خلاصه‌سازی سنتی، از الگوریتم MaxWorth استفاده می‌کنیم و متنی را انتخاب می‌کنیم که بیشترین میانگین ارزش سنجش را دارد. برای اعتبارسنجی این انتخاب، ما یک بار متن را با استفاده از روش TF-IDF[2] خلاصه کردیم. TF-IDF (فرکانس واژه– معکوس فرکانس سند) یک تکنیک برای خلاصه سازی متن با استفاده از ارزیابی اهمیت کلمات در یک مجموعه اسناد است. این تکنیک اهمیت یک کلمه را در یک سند خاص نسبت به فراوانی آن در کل مجموعه سندها اندازه‌گیری می‌کند. عنصر فرکانس واژه، فراوانی یک واژه در یک سند را ارزیابی می‌کند به این صورت که وقوع‌های بیشتر نشان‌دهنده اهمیت بیشتر واژه است. اما برای متعادل کردن اهمیت واژگانی که به طور متداول در بسیاری از اسناد موجودند، عنصر معکوس فرکانس سند وزن این واژه‌ها را کاهش می‌دهد. این روش نرمال‌سازی، امکان استخراج واژگان کلیدی که منحصر به فرد برای یک سند هستند و اهمیت معنایی بیشتری دارند را فراهم می‌کند. جملاتی که دربرگیرنده کلمات با امتیاز TF-IDF بیشتری هستند، به عنوان مهم‌ترین جملات مرتبط با محتوای سند در نظر گرفته می‌شوند.

یک بار دیگر نیز، اولین ۵۱۲ کلمه متن را بدون هیچ گونه خلاصه‌سازی به BERT وارد کردیم. در هر دوی این حالات مدل را آموزش داده و سنجیدیم. نتایج آزمایش این مدل‌ها در اثر روش انتخاب گسترهٔ متن بر روی کارآیی مدل نشان داده شده است. همانطور که مشاهده می شود، MaxWorth در این موضوع کمی بهتر از سایرین عمل کرده است.

---

[1] Bert-large
[2] Term Frequency – Inverse Document Frequency

جدول ٦ - اثر روش انتخاب گسترهٔ متن بر روی کارآیی مدل

روش انتخاب متن	صحت	دقت	یادآوری	امتیاز F1	AUC
روش MaxWorth	۰.۸۵۴	۰.۷۵۶	۰.۵۵۵	۰.۶۴۰	۰.۸۶۴
روش TFIDF	۰.۸۱۰	۰.۵۷۹	۰.۶۹۳	۰.۶۳۱	۰.۸۴۹
اولین ۵۱۲ کلمه	۰.۸۴۷	۰.۸۴۶	۰.۴۲۲	۰.۵۶۰	۰.۸۴۸

۵-۱-۲- معماری شبکه خروجی

در بخش ۴-۲، استدلال کردیم که معماری شبکه خروجی باید تا حد امکان ساده باشد تا مکانیزم خود توجهی BERT بتواند به خوبی کار خود را انجام دهد. برای بررسی این ادعا، ما سه مدل مبتنی بر BERT را با شبکه های خروجی مختلف مقایسه کردیم. برای انصاف در مقایسه، به عنوان یک مدل با شبکه خروجی ساده، از مدلی بدون انتخاب متن واجد ارزش سنجش استفاده کردیم که تنها یک لایه BERT برای عنوان و متن داشت و به دنبال آن یک خروجی و یک شبکه عصبی خطی وجود داشت. اولین مدل با شبکهٔ خروجی پیچیده همان مدل FakeBERT است که قبلا معرفی کردیم. همانطور که قبلا ذکر شد، FakeBERT دارای یک شبکه خروجی پیچیده، شامل سه شبکه عصبی کانولوشن و بسیاری از لایه‌های کمکی دیگر است. در یک طراحی معماری دیگر، دو لایه حافظهٔ کوتاه‌مدت طولانی[1] و دو لایه حذف تصادفی را به شبکه خروجی مدل خطی که در بالا اشاره شد، اضافه کردیم. یک بار هم به جای حافظه کوتاه‌مدت طولانی معمولی از حافظه کوتاه‌مدت طولانی دوطرفه[2] استفاده کردیم. همانطور که نتایج آزمایش‌ها را می توان در جدول ۷ مشاهده کرد، شبکه های خروجی پیچیده چیزی به این مدل های مبتنی بر BERT اضافه نمی‌کنند.

جدول ۷- اثر معماری شبکه خروجی بر کارآیی مدل

شبکه خروجی (مدل)	صحت	دقت	بازیابی	امتیاز F1	AUC
کانولوشن پیچیده (FakeBERT)	۰.۸۰۱	۰.۵۶۰	۰.۷۰۵	۰.۶۲۴	۰.۸۵۴
حافظهٔ کوتاه مدت طولانی یک‌طرفه (StanceBert)	۰.۸۴۵	۰.۷۰۳	۰.۵۸۳	۰.۶۳۷	۰.۸۴۰
حافظهٔ کوتاه مدت طولانی دوطرفه (StanceBert)	۰.۷۷۳	۰.۷۲۳	۰.۵۰۰	۰.۵۹۱	۰.۸۲۳
خطی (StanceBert)	۰.۸۵۴	۰.۷۵۶	۰.۵۵۵	۰.۶۴۰	۰.۸۶۴

[1] Long Short Term Memory
[2] BiLSTM

۵-۱-۳- اندازه ورودی

زمان آموزش با افزایش اندازه ورودی افزایش می‌یابد. بنابراین باید دلایل خوبی برای افزایش اندازه ورودی داشته باشیم. ما تصمیم گرفتیم عملکرد مدل را با تغییر طول ورودی BERT اندازه گیری کنیم تا ببینیم افزایش طول ورودی چگونه گزینه خوبی برای ما خواهد بود. بنابراین، ما مدل BERT بدون موازی‌سازی را با سه اندازه ورودی مختلف آموزش دادیم: ۱۲۸، ۲۵۶، و ۵۱۲. همانطور که در جدول ۸ مشاهده می شود، افزایش اندازه ورودی عملکرد مدل را اندکی بهبود می‌بخشد. با این حال، اندازه های ورودی بزرگتر به حافظه بیشتری نیاز دارند و به دلیل امکانات محاسباتی ما نمی‌توانیم اندازه ورودی را به مقدار دلخواه افزایش دهیم.

جدول ۸- اثر اندازه ورودی بر کارآیی مدل

اندازه ورودی	صحت	دقت	یادآوری	امتیاز F1	AUC
۱۲۸	۰.۷۵۱	۰.۴۸۰	۰.۷۷۰	۰.۵۹۲	۰.۸۵۰
۲۵۶	۰.۸۱۱	۰.۵۸۰	۰.۶۹۸	۰.۶۳۴	۰.۸۵۷
۵۱۲	۰.۸۴۷	۰.۸۴۶	۰.۴۲۲	۰.۵۶۳	۰.۸۴۸

٤-۱-۵- استفاده از معماری‌های دیگر برت

توسعه‌دهندگان برت علاوه بر مدل پایه برت، مدل بزرگتری را نیز پیش‌آموزش داده‌اند که دارای ۲۴ لایهٔ انتقال‌دهنده به جای ۱۲ لایهٔ مدل برت پایه است. ما در مدل خود یک بار هم از مدل بزرگ برت برای کدگذاری متن خبر استفاده کردیم و نتایج را با مدل اصلی خود سنجیدیم که این مقایسه در جدول ۹ آمده است. همان‌طور که مشاهده می‌شود، استفاده از مدل بزرگتر برت نه تنها باعث افزایش کارآیی نشده بلکه آن را کاهش نیز داده است. به نظر می‌رسد، دلیل این امر کم بودن تعداد نمونه‌های موجود است که نمی‌تواند از عهده ریزتنظیم چنین مدل بزرگی برآید.

جدول ۹ - تاثیر استفاده از مدل برت بزرگ

شبکه برت مورد استفاده	صحت	دقت	یادآوری	امتیاز F1	AUC
Bert-base	۰.۸۵۴	۰.۷۵۶	۰.۵۵۵	۰.۶۴۰	۰.۸۶۴
Bert-large	۰.۷۵۰	۰.۷۲۳	۰.۴۹۵	۰.۵۸۸	۰.۸۰۱

۶- نتیجه‌گیری

تشخیص اخبار جعلی نه تنها می‌تواند به مقابله با انتشار اطلاعات نادرست کمک کند، بلکه به‌عنوان یکی از ابزارهای کلیدی در مقابله با جنگ سایبری و تقویت پدافند غیرعامل عمل می‌کند. با توجه به نقش مهم اخبار جعلی در ایجاد تفرقه، کاهش اعتماد اجتماعی و تضعیف منابع رسمی، توسعه این تکنیک‌ها می‌تواند به بهبود امنیت ملی و اطلاعاتی کمک کند. تشخیص اخبار جعلی یک نوع طبقه‌بندی متن است که به دلیل ویژگی‌های خاص خود، علاوه بر متون عمومی، دارای ابعاد منحصربه‌فردی است. به عنوان مثال، اخبار معمولاً طولانی هستند و همه بخش‌های یک خبر به یک اندازه در تشخیص صحت آن تأثیرگذار نیستند؛ همچنین، هر خبر دارای یک تیتر است. عموماً در خبرهای جعلی رابطهٔ استنباطی قوی‌ای بین عنوان خبر و متن آن وجود ندارد. بر این اساس، سعی کردیم تا مدلی طراحی کنیم که بتواند موضع عنوان خبر و متن آن را نیست به هم تشخیص دهد. برای این کار، ابتدا عنوان خبر و مهم‌ترین بخش‌های متن خبر را با استفاده از الگوریتم MaxWorth انتخاب کردیم. سپس، این داده‌ها را با استفاده از دو شبکه BERT موازی کدگذاری کرده و به یک طبقه‌بند خطی وارد کردیم. نتایج حاکی از آن بود که این انتخاب‌ها منجر به بهبود عملکرد مدل شده است.

در بررسی نتایج آزمایش‌ها، مشخص شد که استفاده از الگوریتم MaxWorth برای انتخاب بخش‌هایی از مقالات طولانی به عنوان ورودی مدل، نسبت به الگوریتم‌های خلاصه‌سازی بهتر عمل می‌کند. همچنین، به‌کارگیری دو شبکه BERT موازی برای تحلیل موضع بین تیتر و بدنه خبر، رویکرد موفقی بود و کارایی بهتری نسبت به استفاده از یک شبکه BERT به‌تنهایی داشت. تجربه نشان داد که استفاده از شبکه‌های پیچیده‌تر پس از BERT، نه تنها محاسبات را سنگین‌تر می‌کند، بلکه تأثیر مثبت وزن‌های BERT را کاهش می‌دهد و کارایی مدل را تحت‌الشعاع قرار می‌دهد. بنابراین، بهتر است از یک شبکه عصبی ساده خطی برای خروجی استفاده شود تا مدل بتواند به شکل بهینه‌تری خروجی BERT را به خروجی نهایی تبدیل کند. همچنین دیده شد که افزایش طول ورودی به مدل BERT می‌تواند بهبود جزئی در کارایی مدل ایجاد کند.

برای بهبود بیشتر در تشخیص اخبار جعلی با استفاده از شبکه‌های عمیق، نیاز مبرم به داده‌های بیشتر داریم. با در اختیار داشتن مجموعه داده‌های بزرگ‌تر، امکان استفاده از مدل‌های تبدیل‌کنندهٔ بزرگ‌تر و پیچیده‌تر فراهم می‌شود که به دلیل پارامترهای بیشتر، می‌توانند به شکل بهتری یادگیری داشته باشند. یکی از اولویت‌های مهم در آینده، جمع‌آوری هر چه بیشتر داده‌های مربوط به اخبار جعلی است. همچنین، برخی از مدل‌های پیش‌آموزش‌یافته BERT به صورت چندزبانه طراحی شده‌اند و این امکان را فراهم می‌کنند که از وزن‌های یادگرفته شده در یک زبان، در زبان‌های دیگر نیز

استفاده کنیم. این رویکرد، نه تنها جمع‌آوری داده‌ها از زبان‌های مختلف را تسهیل می‌کند، بلکه ابزاری برای ارزیابی اخبار در زبان‌هایی که منابع کمی دارند نیز فراهم می‌سازد.

علاوه بر این، یکی از برنامه‌های آینده ما این است که از منابع مختلف خبری برای تشخیص جعلی بودن اخبار استفاده کنیم. این منابع شامل نظرات و بازخوردهای کاربران در شبکه‌های اجتماعی، قالب‌های غیرمتنی مانند عکس‌ها، ویدیوها و فایل‌های صوتی، و همچنین نحوه انتشار اخبار در شبکه‌های اجتماعی می‌شوند.

۷–تشکر و قدردانی

در اینجا جا دارد از جناب آقای دکتر وحید رافع که با راهنمایی‌های خود اینجانب را در انجام این پروژه یاری رساندند تقدیر و تشکر به عمل آورم.

۸–تعارض منافع

نویسنده(گان) اعلام می‌دارند که در مورد انتشار این مقاله تضاد منافع وجود ندارد. علاوه بر این، موضوعات اخلاقی شامل سرقت ادبی، رضایت آگاهانه، سوء رفتار، جعل داده‌ها، انتشار و ارسال مجدد و مکرر توسط نویسندگان رعایت شده است.

۹–دسترسی آزاد

۱۰- منابع

[1]. Meel P, Vishwakarma DK (2020) Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Syst Appl 153:112986
[2]. Kwon S, Cha M, Jung K, Chen W, Wang Y (2013) Prominent features of rumor propagation in online social media. In: Proceedings - IEEE International Conference on Data Mining, ICDM. pp 1103–1108
[3]. Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake News Detection on Social Media. ACM SIGKDD Explorations Newsletter 19:22–36
[4]. Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science (1979) 359:1146–1151
[5]. Karimi H, Tang J (2019) Learning hierarchical discourse-level structure for fake news detection. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference 1:3432–3442
[6]. Thorne J, Vlachos A, Christodoulopoulos C, Mittal A (2018) FEVER: A large-scale dataset for fact extraction and verification. In: NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference. pp 809–819
[7]. Bronstein MM, Bruna J, Lecun Y, Szlam A, Vandergheynst P (2017) Geometric Deep Learning: Going beyond Euclidean data. IEEE Signal Process Mag 34:18–42
[8]. Monti F, Frasca F, Eynard D, Mannion D, Bronstein MM (2019) Fake News Detection on Social Media using Geometric Deep Learning. Arxiv 1902-06673v1 1–15
9. Albawi S, Mohammed TA, Al-Zawi S (2018) Understanding of a convolutional neural network. In: Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017. pp 1–6
[10]. Zhou X, Zafarani R (2020) A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. ACM Comput Surv. https://doi.org/10.1145/3395046
[11]. Wang B, Feng Y, Xiong X cai, Wang Y heng, Qiang B hua (2022) Multi-modal transformer using two-level visual features for fake news detection. Applied Intelligence 1–15
[12]. Wu K, Yang S, Zhu KQ (2015) False rumors detection on Sina Weibo by propagation structures. In: Proc Int Conf Data Eng. pp 651–662
[13]. Shu K, Wang S, Liu H (2019) Beyond news contents: The role of social context for fake news detection. In: WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining. pp 312–320
[14]. Zhou X, Cao J, Jin Z, Xie F, Su Y, Zhang J, Chu D, Cao X (2015) Real-time news certification system on sina weibo. In: WWW 2015 Companion - Proceedings of the 24th International Conference on World Wide Web. pp 983–988
[15]. Pan JZ, Pavlova S, Li C, Li N, Li Y, Liu J (2018) Content based fake news detection using knowledge graphs. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in

Bioinformatics). Springer, pp 669–683
[16]. Ott M, Choi Y, Cardie C, Hancock JT (2011) Finding deceptive opinion spam by any stretch of the imagination. In: ACL-HLT 2011 - Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. pp 309–319
[17]. Abouelenien M, Pérez-Rosas V, Zhao B, Mihalcea R, Burzo M (2017) Gender-based multimodal deception detection. In: Proceedings of the ACM Symposium on Applied Computing. Association for Computing Machinery, New York, New York, USA, pp 137–144
[18]. Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R (2018) Automatic detection of fake news. COLING 2018 - 27th International Conference on Computational Linguistics, Proceedings 2789:3391–3401
[19]. Gravanis G, Vakali A, Diamantaras K, Karadais P (2019) Behind the cues: A benchmarking study for fake news detection. Expert Syst Appl 128:201–213
[20]. Alghamdi J, Lin Y, Luo S (2024) Unveiling the hidden patterns: A novel semantic deep learning approach to fake news detection on social media. Eng Appl Artif Intell 137:109240
[21]. Pierri F, Ceri S (2019) False news on social media: A data-driven survey. SIGMOD Rec 48:18–32
[22]. Panigrahi S, Nanda A, Swarnkar T (2021) A Survey on Transfer Learning. Smart Innovation, Systems and Technologies 194:781–789
[23]. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations ofwords and phrases and their compositionality. Adv Neural Inf Process Syst
[24]. Pennington J, Socher R, Manning CD (2014) GloVe: Global vectors for word representation. In: EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference. pp 1532–1543
[25]. Peters ME, Neumann M, Zettlemoyer L, Yih WT (2018) Dissecting contextual word embeddings: Architecture and representation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018. pp 1499–1509
[26]. Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference 1:4171–4186
[27]. Luong MT, Pham H, Manning CD (2015) Effective approaches to attention-based neural machine translation. In: Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing. pp 1412–1421
[28]. Rajpurkar P, Zhang J, Lopyrev K, Liang P (2016) SQuad: 100,000+ questions for machine comprehension of text. In: EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings. pp 2383–2392
[29]. Williams A, Nangia N, Bowman SR (2018) A broad-coverage challenge corpus for sentence understanding through inference. In: NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference. pp 1112–1122

[30]. Shu K, Mahudeswaran D, Wang S, Lee D, Liu H (2020) FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. Big Data 8:171–188
[31]. Singhania S, Fernandez N, Rao S (2017) 3HAN: A Deep Neural Network for Fake News Detection. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10635 LNCS:572–581
[32]. Kaliyar RK, Goswami A, Narang P (2021) FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimed Tools Appl 80:11765–11788
[33]. Loshchilov I, Hutter F (2019) Decoupled weight decay regularization. 7th International Conference on Learning Representations, ICLR 2019